Error Estimation in Krylov Subspace Methods for Matrix Functions

Andreas Frommer, **Marcel Schweitzer**

Bergische Universität Wuppertal

October 26, 2015

Outline

Krylov subspace approximation to $f(A)\boldsymbol{b}$

Stieltjes functions

Computation of Lanczos error bounds
- First step: Error representation
- Second step: Error norm as quadratic form
- Third step: Gauss quadrature
- Fourth step: Getting things efficient

Numerical examples

Conclusions & Outlook

Matrix functions

Definition: Matrix function

Let $A \in \mathbb{C}^{n \times n}$, $\operatorname{spec}(A) \subset \Omega \subset \mathbb{C}$ and $f : \Omega \longrightarrow \mathbb{C}$ sufficiently smooth. Then define

$$f(A) := p(A)$$

where p is the polynomial that interpolates f at $\operatorname{spec}(A)$ in the Hermite sense.

- **Wanted:** $f(A)\boldsymbol{b}$, the action of $f(A)$ on a vector $\boldsymbol{b} \in \mathbb{C}^n$.
- **Problem:** $f(A)$ is full, even when A is sparse.
- $\rightsquigarrow$ Forming $f(A)$ explicitly and then multiplying it to $\boldsymbol{b}$ is not feasible for large A.

Krylov subspace approximation to $f(A)\boldsymbol{b}$

- **Therefore:** Try to approximate $f(A)\boldsymbol{b}$ directly.

Definition: Krylov subspace

Let $A \in \mathbb{C}^{n\times n}, \boldsymbol{b} \in \mathbb{C}^n$. Then the mth *Krylov subspace* with respect to A and $\boldsymbol{b}$ is

$$\mathcal{K}_m(A,\boldsymbol{b}) = \{p_m(A)\boldsymbol{b} : p_m \text{ polynomial with} \deg p_m < m\}.$$

- As $f(A)\boldsymbol{b} = p(A)\boldsymbol{b}$, approximate

$$f(A)\boldsymbol{b} \approx \boldsymbol{f}_m \in \mathcal{K}_m(A,\boldsymbol{b}).$$

Krylov subspace approximation to $f(A)\boldsymbol{b}$

- Let A be Hpd, $V_m = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_m]$ the orthonormal basis obtained from m steps of the Lanczos method, then

$$AV_m = V_m T_m + t_{m+1,m} \boldsymbol{v}_{m+1} \boldsymbol{e}_m^H$$

with a tridiagonal matrix T_m.

- Define mth *Lanczos approximation* to $f(A)\boldsymbol{b}$ as

$$\boldsymbol{f}_m = \|\boldsymbol{b}\|_2 V_m f(T_m) \boldsymbol{e}_1.$$

- **Question:** How well does $\boldsymbol{f}_m$ approximate $f(A)\boldsymbol{b}$, i.e., what do we know about

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|_2 \text{ ?}$$

$\rightsquigarrow$ When can we stop the iteration?

Stieltjes functions

Definition: Stieltjes function

A function $f : \mathbb{C} \setminus \mathbb{R}_0^- \longrightarrow \mathbb{C}$ defined by

$$f(z) = \int_0^\infty \frac{1}{z+t} d\mu(t)$$

with a non-negative, monotonically increasing function μ is called *Stieltjes function*.

Lemma

If $\mu(t)$ is differentiable, then

$$\int_0^\infty \frac{1}{z+t} d\mu(t) = \int_0^\infty \frac{\mu'(t)}{z+t} \, \mathrm{d}t.$$

Stieltjes functions

Examples of Stieltjes functions

1. $f(z) = \frac{1}{z}$
2. $f(z) = \sum_{i=1}^{\ell} \frac{a_i}{z+b_i}, a_i, b_i \geq 0$
3. $f(z) = z^{-\alpha}, \alpha \in (0,1)$
4. $f(z) = \frac{1-e^{-\theta\sqrt{z}}}{z}$
5. $f(z) = \frac{\log(1+z)}{z}$

- **Applications:** Solution of PDEs, sampling from Gaussian Markov random fields, Dirichlet-to-Neumann maps, quantum chromodynamics, ...

Properties of Stieltjes functions

- For f a Stieltjes function, $f(z) \in \mathbb{R}^+$ when $z \in \mathbb{R}^+$
- Derivative of Stieltjes functions

Lemma

Let f be a Stieltjes function. Then

$$f^{(j)}(z) = (-1)^{j+1} \int_0^\infty \frac{1}{(z+t)^{j+1}} \, \mathrm{d}\mu(t) \text{ for all } k \in \mathbb{N}$$

- Stieltjes functions are completely monotonic on $\mathbb{R}^+$, i.e.,

$$(-1)^j f^{(j)}(z) \geq 0 \text{ for all } j = 0, 1, 2, \ldots \text{ and } z \in \mathbb{R}^+.$$

First step: Error representation

Using the connection to polynomial interpolation, we find

Theorem [Frommer, Güttel, S 2014]

Let f be a Stieltjes function and let $\boldsymbol{f}_m$ be the mth Lanczos approximation to $f(A)\boldsymbol{b}$. Let $\text{spec}(T_m) = \{\theta_1, \ldots, \theta_m\}$ and define

$$e_m(z) = (-1)^{m+1}\|\boldsymbol{b}\|_2\gamma_m \int_0^\infty \frac{1}{w_m(t)} \cdot \frac{1}{z+t}\,\mathrm{d}\mu(t),$$

where $w_m(t) = (t+\theta_1)\cdots(t+\theta_m)$ and $\gamma_m = \prod_{i=1}^m t_{i+1,i}$. Then

$$f(A)\boldsymbol{b} - \boldsymbol{f}_m = e_m(A)\boldsymbol{v}_{m+1},$$

where $\boldsymbol{v}_{m+1}$ is the $(m+1)$st Lanczos vector.

First step: Error representation

- Difference between original function and error function: *Reciprocal nodal polynomial* $1/w_m(t)$ in integrand
- A Hpd $\Rightarrow$ all Ritz values real & positive
- $1/w_m(t)$ positive and monotonically decreasing
- $\widetilde{\mu}(t) = \int_0^t \frac{1}{w_m(\tau)}\, \mathrm{d}\mu(\tau)$ positive, monotonically increasing and bounded.

First step: Error representation

- Difference between original function and error function: *Reciprocal nodal polynomial* $1/w_m(t)$ in integrand
- A Hpd $\Rightarrow$ all Ritz values real & positive
- $1/w_m(t)$ positive and monotonically decreasing
- $\widetilde{\mu}(t) = \int_0^t \frac{1}{w_m(\tau)} \,\mathrm{d}\mu(\tau)$ positive, monotonically increasing and bounded.

Lemma [Frommer, Güttel, S 2014]

When A is Hpd, the error function can be written as

$$e_m(z) = (-1)^{m+1} \|\boldsymbol{b}\|_2 \gamma_m \widetilde{e}_m(z),$$

where

$$\widetilde{e}_m(z) = \int_0^\infty \frac{1}{t+z} \,\mathrm{d}\widetilde{\mu}(t) \text{ with } \mathrm{d}\widetilde{\mu}(t) = \frac{1}{w_m(t)} \,\mathrm{d}\mu(t).$$

is a Stieltjes function.

Second step: Error norm as quadratic form

- $f(A)\boldsymbol{b} - \boldsymbol{f}_m = (-1)^{m+1}\|\boldsymbol{b}\|_2\gamma_m\widetilde{e}_m(A)\boldsymbol{b}$ implies

Corollary [Frommer, S 2015]

Let A be Hpd, then

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|_2^2 = \|\boldsymbol{b}\|_2^2\gamma_m^2\boldsymbol{v}_{m+1}^H\widetilde{e}_m(A)^2\boldsymbol{v}_{m+1}.$$

- Straightforward, but helpful result:

Proposition

Let f, g be completely monotonic on $\mathbb{R}^+$. Then $f \cdot g$ is completely monotonic on $\mathbb{R}^+$.

$\Rightarrow \widetilde{e}_m(z)^2$ is completely monotonic.

Third step: Gauss quadrature

- **So far:** Lanczos error norm for Stieltjes function is a quadratic form induced by a completely monotonic function
- How can we approximate such a quantity?
- Make things more complicated: **This talk needs more integrals!**
- Let $A = Q\Lambda Q^H$, $\eta = Q^H \boldsymbol{v}_{m+1}$. Then

$$\boldsymbol{v}_{m+1}^H \widetilde{e}_m(A)^2 \boldsymbol{v}_{m+1} = \int_{\lambda_1}^{\lambda_n} \widetilde{e}_m(z)^2 \, \mathrm{d}\alpha(z)$$

with

$$\alpha(z) = \begin{cases} 0 & z \leq \lambda_1 \\ \sum_{j=1}^{i} |\eta_j|^2 & \lambda_i < z \leq \lambda_{i+1} \\ \sum_{j=1}^{n} |\eta_j|^2 & \lambda_n < z \end{cases}$$

Third step: Gauss quadrature

- Lanczos error norm is a Riemann–Stieltjes integral of a completely monotonic function
- Gauss quadrature gives bounds for such integrals

Theorem

Let g be completely monotonic on $[\lambda_1, \lambda_n]$ and let z_ℓ, ω_ℓ and $\widetilde{z}_\ell, \widetilde{\omega}_\ell$ be the nodes and weights of the k-point Gauss and $(k+1)$-point Gauss–Radau quadrature rule (with one node fixed at λ_1) for $\int_{\lambda_1}^{\lambda_n} g(z)\,\mathrm{d}\alpha(z)$, respectively. Then

$$\sum_{\ell=1}^{k} \omega_\ell g(z_\ell) \leq \int_{\lambda_1}^{\lambda_n} g(z)\,\mathrm{d}\alpha(z)$$

and

$$\sum_{\ell=1}^{k+1} \widetilde{\omega}_\ell g(\widetilde{z}_\ell) \geq \int_{\lambda_1}^{\lambda_n} g(z)\,\mathrm{d}\alpha(z).$$

Third step: Gauss quadrature

- So, things are **easy!**
- Determine error function and use Gauss quadrature to find bounds for the Lanczos error norm
- **Oh, wait...** We don't even know α explicitly

Third step: Gauss quadrature

- So, things are **easy!**
- Determine error function and use Gauss quadrature to find bounds for the Lanczos error norm
- **Oh, wait...** We don't even know α explicitly
- The "Matrices, moments & quadrature" idea helps:

Theorem [Golub, Meurant]

Let $A \in \mathbb{C}^{n \times n}$ be Hpd, $\boldsymbol{v}_{m+1} \in \mathbb{C}^n$. Let z_ℓ, ω_ℓ be the nodes and weights of the k-point Gaussian quadrature rule for approximating $\int_{\lambda_1}^{\lambda_n} g(z)\,\mathrm{d}\alpha(z)$ with α as before. Then

$$\sum_{\ell=1}^{k} \omega_\ell g(z_\ell) = \boldsymbol{e}_1^H g(T_k^{(2)}) \boldsymbol{e}_1,$$

$T_k^{(2)}$ the tridiag. matrix from k Lanczos steps for A and $\boldsymbol{v}_{m+1}$.

Third step: Gauss quadrature

- Putting everything together gives

Theorem [Frommer, S 2015]

Let f be a Stieltjes function, let $A \in \mathbb{C}^{n \times n}$ be Hpd, let $\boldsymbol{f}_m$ be the mth Lanczos approximation to $f(A)\boldsymbol{b}$. Let $T_k^{(2)}$ be the tridiagonal matrix resulting from k steps of the Lanczos process for A and $\boldsymbol{v}_{m+1}$ and let $\widetilde{T}_k^{(2)}$ be the modification of $T_k^{(2)}$ corresponding to Gauss–Radau quadrature. Then

$$\|\boldsymbol{b}\|_2^2 \gamma_m^2 \boldsymbol{e}_1^H \widetilde{e}_m \left(T_k^{(2)}\right)^2 e_1 \leq \|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|_2^2 \leq \|\boldsymbol{b}\|_2^2 \gamma_m^2 \boldsymbol{e}_1^H \widetilde{e}_m \left(\widetilde{T}_k^{(2)}\right)^2 e_1.$$

- So, we do Lanczos to approximate the error in Lanczos...
  Sure... sounds like an awesome plan!

Fourth step: Getting things efficient

- Naive computation of k-node error bounds:
  $\rightsquigarrow k$ additional multiplications with A
- First thought: It's probably better to use these multiplications for our main Lanczos method and get a more accurate iterate...
- **Or maybe we just use them for both?!**

Fourth step: Getting things efficient

- Naive computation of k-node error bounds:
  $\rightsquigarrow k$ additional multiplications with A
- First thought: It's probably better to use these multiplications for our main Lanczos method and get a more accurate iterate...
- **Or maybe we just use them for both?!**
- Lanczos restart recovery:

Theorem [Frommer, Kahl, Lippert, Rittich 2013]

Let $\hat{T}$ be the bottom right $(2k+1) \times (2k+1)$ submatrix of T_{m+k+1}. Then k steps of the Lanczos process for $\hat{T}$ and $\boldsymbol{e}_{k+1}$ produce the same matrix $T_k^{(2)}$ as k steps of the Lanczos process for A and $\boldsymbol{v}_{m+1}$.

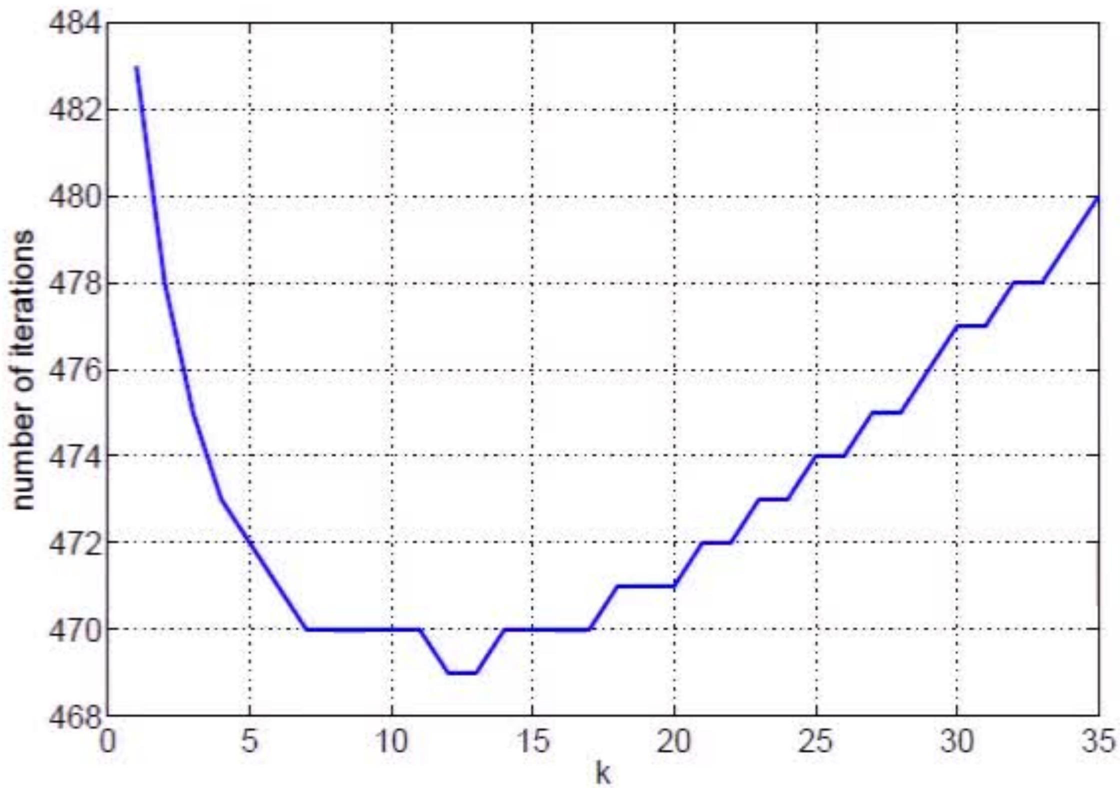
Fourth step: Getting things efficient

- So what exactly does restart recovery give us now?

- All multiplications with A advance the primary Lanczos
- We can still recover everything we need to compute error bounds (with additional cost $\mathcal{O}(k^2)$)
- But: Error bounds for iterate from step m are not available before step $m+k+1$
- Lanczos converges monotonically for A Hpd $\rightsquigarrow$ error bounds computed for $\boldsymbol{f}_m$ are also valid for $\boldsymbol{f}_{m+k+1}$.

Example: Lattice QCD

- Trade-off between accuracy of the bounds and timely availability.
- Iteration number in which the upper bound is below 10^{-9}:



Conclusions & Outlook

Conclusions (+ things I didn't talk about):

- For $A \in \mathbb{C}^{n \times n}$ Hpd, guaranteed error bounds for the mth Lanczos approximation $\boldsymbol{f}_m$ to $f(A)\boldsymbol{b}$ can be computed with cost independent of m and n
- The same techniques can be used to compute estimates (but no bounds in general) in the non-Hermitian case with cost independent of n
- Similar results for extended/rational Krylov subspace methods
- Several variants of (rational) restart recovery for these subspaces

Outlook

- Theoretical analysis for predicting quality of the bounds